

University of Groningen

Logics of scientific cognition. Reply to Johan van Benthem

Kuipers, Theo A.F.

Published in:
EPRINTS-BOOK-TITLE

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2005

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kuipers, T. A. F. (2005). Logics of scientific cognition. Reply to Johan van Benthem. In *EPRINTS-BOOK-TITLE* University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Theo A. F. Kuipers

LOGICS OF SCIENTIFIC COGNITION

REPLY TO JOHAN VAN BENTHEM

In brainstorming about a suitable title for the present book the editors pondered over “Logics of Scientific Cognition.” For some reason this title did not make it, but Johan van Benthem’s contribution would have fitted that description perfectly. It breathes the same flexible spirit that was already present in his review-like programmatic paper “The Logical Study of Science” (Van Benthem 1982) and that seems to me essential for the very idea of confronting logic and philosophy of science. As also mentioned in my reply to Batens, unfortunately for philosophy of science, after 1982, with a few exceptions (e.g. Van Benthem 1987), Van Benthem directed his logical skills mainly to a number of other areas of the application of logic, notably informatics, linguistics and the philosophy of nature, in particular the logic of time. It is to be hoped that his present contribution is a sign of a partial return to philosophy of science.

In this reply to Van Benthem I start with a number of miscellaneous remarks that cropped up when reading his Sections 2 and 3.2. I shall then indicate that the notion of “partial theory” propagated by him in Section 3.3 not only seems to be directly useful for a couple of interesting topics in the philosophy of science. It has also some resemblance to an interesting special type of theory that I have called “partition theory” (Kuipers 1982). One reason for not dealing with Van Benthem’s Sections 4 and 5 is to challenge him, his colleagues and his students to really cross the threshold and to reside for a substantial time in the philosophy of science. There is really much interesting work to do for the mutual benefit of philosophy of science and logic. For this general theme, see also a few other replies in this volume, notably to Aliseda, Meheus, Batens, Zwart, Mormann, and Burger and Heidema, and in the companion volume those to Kamps and Ruttkamp.

In: R. Festa, A. Aliseda and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth Approximation (Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 83)*, pp. 420-427. Amsterdam/New York, NY: Rodopi, 2005.

Miscellaneous Remarks

On Section 2.1: Theory Formats

(1) Van Benthem states that in ordinary reasoning theorizing about (using the terms of my reply to Sjoerd Zwart) *one* intended application and *a class* of intended applications may play together. I would like to emphasize that precisely the same applies in the theory-experiment interplay in science (ICR, Section 7.3, briefly indicated in Van Benthem's Section 3.2). In experimental contexts we try to establish by "actual truth approximation," given a vocabulary, the true description (or theory, if you like) of one experiment, that is, its set-up, initial conditions, and the outcome. By "nomic truth approximation" we try to approach the strongest true theory of what is physically possible, that is, the set of physical or nomic possibilities, using the results of (attempts at) actual truth approximation with regard to experiments, including inductive leaps based on them.

(2) Van Benthem rightly notes that by defining theories as sets of models, and, I should add, hence also by conceiving theories as deductively closed sets of statements, one "loses the *packaging* into chunks in which the information was presented." The modern definitions of theories he is hinting at by $\text{MOD}^+(T)$ may be very useful for localizing possible causes for counterexamples to theories, and hence for theory revision in favor of closer-to-the-truth ones in the basic sense of ICR. This is *a fortiori* the case when information chunks are ordered. By quoting from SiS, I would like to stress that in the philosophy of science very nice, partially computational, work in this direction has already been done:

Finally, Darden [1990] considers the solution of an anomaly of a theory as a task for diagnostic reasoning, which has been developed for expert systems, guided by the tracing of a defect in a technical system. She shows, using Mendelian examples (see also Darden (1991)), that the decomposition of all presuppositions of a theory may provide the points of departure for solving the anomaly, and that its solution, as in Lakatos' famous analysis (1976), may or may not lead to fundamental theory revision, that is, a revision staying within a research program or breaking through the barriers of the relevant program. (SiS, p. 301).

On Section 2.2: Information Update

(3) I shall return later to the claim that "information update proceeds by elimination of possibilities." Here I just want to note that the party illustration of "elimination of possibilities," in the sense of possible models, typically amounts to actual truth approximation by applying *given* general facts, where the latter happen to be, in conjunction, so strong that they leave only room for one final possibility. That is, relative to the given language, an incomplete true

description is extended to the complete true one by eliminating all the other ones. By leaving out at least one such general fact and adding an initial condition, we obtain formal analogues of actual truth approximation as it occurs in the HD prediction and DN explanation of individual facts. These types of actual truth approximation are essentially contained in the general approach to propositional languages presented in (Kuipers 1982, Section 4). This general approach also covers another special case: actual truth approximation by successive corrections of a provisional description of a certain situation, whether that situation is realized by nature or by experiment. Only this special case is also presented in ICR (Section 7.3.1).

Moreover, they differ from “nomic truth approximation” by the interplay of realized possibilities and established empirical laws. The crucial reason for the latter is that in all variants of actual truth approximation there is, by definition, one target model, whereas a set of models is the target in the case of nomic truth approximation. More generally, the target is a set in all cases in which we want to establish a certain constellation of possibilities, which happens to be interesting for one reason or another. To mention just one social science example, consider an anthropologist who wants to establish in some way or other “the true moral theory” holding in some society and who has as his target the largest set of “morally allowed actions,” of which we may hope that it is far greater than one. Note that all mentioned types of truth approximation can formally be illustrated by playing with the party example as well as by my favorite electric circuit example.

(4) Regarding “*many-agent* dynamics of update in communication” I just quote a passage from SiS:

Moreover, Thagard [1988] suggests that we consider and simulate a scientific community as a collection of interacting parallel computational systems. Here the idea of social adequacy comes into the picture. The suggested analogy raises the question of the possibility and the desirability of a further division of labor than the well-known one between theoreticians and experimenters. In this connection he thinks in particular of the division of labor between ‘constructive dogmatists’ and ‘skeptical critics’, to use our favorite terms for what is essentially the same idea. (SiS, p. 299).

On Section 3.2: Theories in Progress

(5) By calling R “the result of experimental verifications” van Benthem suggested to me a new explication of the fact that scientists frequently talk about “verification,” where philosophers of science prefer to be more cautious and merely talk about “confirmation.” However, in view of the logical relation between ‘the strongest true theory’ T and one of its models, it is on closer inspection perfectly reasonable to say that the model m that truly characterizes a realized possibility, verifies T , by its definition in terms of nomic

possibilities. Similarly, it verifies any theory X that has m as one of its models. In both cases, just because T and X , respectively, are true on that model. Of course, it would be totally wrong to conclude, for example, that theory X has then been verified in the relevant general sense.

Hence, it might make sense to speak of verification of a theory by a model in contrast to general or overall verification of a theory. Note that this point is directly related to Emma Ruttkamp's observation that theories are frequently overdetermined by the data, instead of underdetermined (see Ruttkamp 2000, her contribution to the companion volume, and my reply).

According to this sense of verification of a theory by a model, an "actualized model" (in the sense of an actualized conceptual possibility) compatible with the theory verifies it. More roughly: a "positive instance" and even a "neutral instance" (see below) of the theory verify it. Hence, a theory can be – at the same time – verified by some models (instances) and falsified by others. This new sense of verification perfectly shows the non-standard way in which I use the term 'model'. See the synopsis' introduction to Part III and my reply to Zwart. The main interest of the new sense of verification may even be restricted to just revealing my non-standard use of the term 'model'.

Note too that the verification of a theory by a model need not imply that we have a confirming instance or model of the theory in the sense of deductive confirmation, discussed in ICR Ch. 2, for the relevant model may be a neutral instance. For example, although both a black raven and non-black non-raven are "verifying and confirming" models of "all ravens are black," a black non-raven is a "verifying but not confirming" model of that hypothesis. This amounts to similar reasons to those in the case of verification in the general sense (as treated in ICR, Ch. 2, pp. 21-2) for not conceiving "verification" simply as an extreme, viz. maximal, form of "confirmation," as is usually done by philosophers of science.

(6) As announced, I come back to van Benthem's claim in 2.2 that "information update proceeds by elimination of possibilities." In 3.2 this is rightly qualified by "we also eliminate potential laws by adding models." Indeed, in addition to my favorite dual formulations of 'more successful' and 'more truthlike' in terms of consequences and models (ICR, p. 182, p. 189), we get in this way also a "dual formulation" of updating the data $R(t)$ and $S(t)$. They are updated to a superset $R(t')$ and subset $S(t')$, respectively, for t' later than t , by eliminating potential laws and possibilities, respectively. Of course, conversely, there is also a dual formulation in terms of "addition" (or "introduction"): adding new possibilities and laws, respectively.

(7) Van Benthem plausibly distinguishes between necessary and sufficient conditions for satisfying, for example, a certain theory. Moreover, he suggests

that the latter form is more appropriate when one concludes that “a whole class of systems – say, all pendulums in a certain range” falls under a physical theory. Although this way of speaking is not false, it may be still more appropriate to conceptualize this as a case of domain change, in particular domain narrowing, and hence change, in particular strengthening, of the relevant truth. For the topic of truth approximation by domain change, see my reply to Sjoerd Zwart.

Partial Theories and Partition Theories

From my point of view two intuitions come together in the notion of a “partial theory.” First, the *three-valued division* of objects, for example of models, into a positive, a negative and a, so far, grey or neutral extension. Second, (*epistemological*) *symmetry* in the sense that “negative instances” are known in the same way as “positive instances.” Both aspects are very important, but they need not go together.

Partition Theories

Let me first turn to “two-valued” epistemically symmetric theories. I have (only) paid special attention to them in my first paper on truth approximation (Kuipers 1982, Section 7.2, pp. 368-371), where I called them “partition theories,” for reasons that will soon become clear. Take an electric circuit with just one bulb and some network of switches. We can immediately assume, for good reasons, that the true theory about the circuit is of the following type: the bulb will light if and only if a certain complex condition for the on and off state of the switches holds. Hence, whereas we may suppose that all conceptually possible switch states are physically possible, they are *partitioned* into two sets, viz. switch states with and without a lit bulb. When we propose a theory for the circuit, we can restrict ourselves to theories of the same form, which hence also partition the set of switch states into two subsets. As a consequence, such a theory can have two types of *real* counterexamples. A lit bulb and a switch state that does not satisfy the theory’s condition and a non-lit bulb and a switch state that does satisfy the condition. The same feature occurs in many other, more interesting, cases. For example, equilibrium theories in physics and economics. The true theory about a balance and hence candidate theories are of the form: the balance is in equilibrium if and only if A quite different example is a grammatical theory. Such a theory is also of the form: a sentence is grammatical, as judged by native speakers, if and only if it satisfies such and such conditions. Hence, a grammatical sentence *not*

satisfying the theory's condition is a counter-example, but an *ungrammatical* sentence *satisfying* the condition is one, too.

In the rest of Section 7.2 of (Kuipers 1982) 'more truthlike' and 'more successful' of partition theories is elaborated, where useful, deviating from the way this is done for theories in general. Taking two types of real counterexamples into account turns out to be less complex than one might expect, even less complex than for theories in general, in which case only one type has to be taken into account. The reason is of course that for theories in general we have also to take a kind of *virtual* counterexamples into account. Recall that $R(t)$ represents the realized, or established *real or nomic*, possibilities at time t and $S(t)$ the, on their basis, strongest established law at t . The complement of $S(t)$ can be said to represent "virtual impossibilities." Hence, whereas $R(t) - X$ represents the established *real* counterexamples of theory X , $X - S(t)$ represents the established *virtual* counterexamples. The asymmetric role attributed to $R(t)$ and $S(t)$ in ICR has everything to do with the special nature of virtual impossibilities and virtual counterexamples. Of course, in some sense, partition theories also have virtual counterexamples. The point is that they are now implicitly represented by real counterexamples. For, to each real counterexample, e.g. a lit bulb with improper switch condition, corresponds a virtual counterexample, e.g. a (non-defective) non-lit bulb in the same switch state.

To be sure, although partition theories frequently occur in the empirical sciences, it is by no means the case that other theories can be missed. On the contrary, in my opinion, the great variety of "empirical laws" implies that as a rule we have to take virtual impossibilities explicitly into account. Only in rather special cases may we have good reasons to suppress them.

Partial Theories

The idea of partial theories, whether or not with a partitioning character in the sense above, does indeed seem to be very useful for philosophy of science purposes. Let me start by indicating how it could refine the general (meta-) theories of confirmation, empirical progress, and truth approximation as presented in ICR. As a matter of fact, ICR and remark (5) above, already contain several indications in this direction. Realized examples of a theory X , that is, members of $R(t) \cap X$, may be of two kinds. X may or may not have some conditional deductive consequence for a member, that is, there may or may not be some partial description of that member for which X entails the rest of the (complete) true description. For confirmation theory, this leads to the distinction between "neutral evidence" (recall the black non-raven for "all ravens are black") and (genuine) "confirming evidence" (a black raven as well as a non-black non-raven), see ICR, e.g. pp. 27-8. Moreover, for the theory of

empirical progress it leads to the symmetric evaluation matrix, according to which, besides negative and positive results, neutral results have to be taken into account when comparing theories (ICR, p. 117). Finally, for the theory of truth approximation and in particular for the definition of ‘more truthlike’ it would lead to a distinction between types of “correct models,” e.g. something like “positive models” and “neutral models,” besides “negative models” of the theory.

As suggested, ICR contains hints with respect to the implementation of the partial character of empirical theories in the first two meta-theories, that is, about confirmation and empirical progress. However, I must confess that I am not sure to what extent I in fact assume in these hints that the empirical theories also are of a partitioning nature. Be this as it may, ICR does not contain hints for implementing “partiality” in the theory of truth approximation. Van Benthem’s proposal at the end of Section 3.3 for this purpose, taking “neutral models” implicitly into account, seems very plausible. The overall “model nature” of that definition suggests that it will nicely combine with a similar overall model formulation of “more successful.” However, such a formulation will very naturally reflect the success comparison of partition theories, by suppressing virtual counterexamples. Hence, an interesting challenge seems to be to give dual formulations for both ‘more successful’ and ‘more truthlike’ of partial theories in general. But I would also not exclude the possibility that, on closer inspection, the dual formulations in ICR of both, which *prima facie* neglect the indicated partial character of theories, already take this character implicitly into account.

I would like to conclude with a straightforward example of partial predicates with a partitioning character. In SiS (Ch. 10) I show that there is a partial analogy between truth approximation and design research. In particular, in the most naïve representation, it is plausible to explicate the idea that a new prototype is an improvement of an old one in view of a set of desired properties (SiS, p. 270) in formally the same “symmetric difference” way as being “more truthlike.” The set of desired or wished for properties W is here supposed to be a subset of a given set RP of “relevant properties.” To be precise, RP does not contain the equally relevant “counterpart properties,” that is, if RP contains P , it does not contain *not-P*. Moreover, $RP-W$ is supposed to represent the set of undesired properties. In this set-up a prototype is partitioning in the sense that ‘not having a desired property’ may be assumed to be equally easy (or difficult) to establish experimentally as ‘having an undesired property’.

One plausible refinement results when one withdraws the assumption that a property that is not desired is automatically undesired. This results in the division of (the potentially relevant) properties into three categories; say,

desired, indifferent and undesired properties, denoted by W^+ , I and W^- . This may be particularly useful in view of later possibilities to negotiate again about W^+ . As long as prototypes are merely described by their properties, the minimal condition of being an improvement is of course that the new prototype has all the desired properties the old one has and the old one has all undesired properties the new one has (SiS, p. 280). However, changing the topic to the market of end products, producers may each have used their own $W^+/I/W^-$ division, and may even have succeeded in realizing products ‘between their W^+ and W^- ’. A consumer will have his own $W^+/I/ W^-$ division. For judging whether he should prefer one of the products over the other, he may well use Van Benthem’s formal definition, at the end of Section 3.3, assuming that he wants to take the intentions of the producers into account. If he does not want to do so, as we may normally expect, he is well advised just to use the minimal condition mentioned above. Hence, our full-blown partial example remains rather artificial, but suggests at the same time that it will be possible to find more realistic examples.

REFERENCES

- Benthem, J. van (1982). The Logical Study of Science. *Synthese* **51**, 431-472.
- Benthem, J. van (1987). Verisimilitude and Conditionals. In: T. Kuipers (ed.), *What is Closer-to-the-Truth?*, pp. 103-128. Amsterdam: Rodopi.
- Darden, L. (1990). Diagnosing and Fixing Faults in Theories. In: J. Shrager and P. Langley (eds.), *Computational Models of Scientific Discovery and Theory Formation*, pp. 319-353. San Mateo: Kaufmann.
- Darden, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. Oxford: Oxford University Press.
- Kuipers, T. (1982). Approaching Descriptive and Theoretical Truth. *Erkenntnis* **18** (3), 343-378.
- Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.
- Ruttkamp, E. (2002). *A Model-Theoretic Realist Interpretation of Science*. Dordrecht: Kluwer Academic Publishers.
- Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge, MA: The MIT Press.